Appendix

Proposal Probabilities for Model Adaptation.

$$
\begin{aligned}
p_{sp}(k,d) &= c^* \cdot \min\{1, \rho/(k+1)\}; \\
p_{me}(k,d) &= c^* \cdot \min\{1, (k-1)/\rho\}; \\
p_{sw}(k,d) &= c^*; \\
& \quad d = 1, \ldots, D; \\
p_{em}(k) &= 1 - \Sigma_{d=1}^{D}[p_{sp}(k,d) + p_{me}(k,d) + p_{sw}(k,d)].
\end{aligned}
$$

Here, c^* is a simulation parameter, and k is the current number of states. The parameter ρ is the hyper-parameter for the truncated Poisson prior probability of the number of states, i.e., ρ is the expected mean of the number of states if the maximum state size is allowed to be $+\infty$, and the scaling factor that multiplies c^* modulates the probability using the resulting state-space size $k \pm 1$ and ρ.

Determining Different Moves in RJ-MCMC

Expectation maximization (EM) is one regular hill-climbing iteration. After a move type other than EM is selected, one or two states at a certain level are selected at random for swap/split/merge, and the parameters are modified accordingly.

Swap the association of two states: Choose two states from the same level, each of which belongs to a different higher-level state; swap their higher-level association.

Split a state: Choose a state at random. The split strategy differs when this state is at different position in the hierarchy: when this is a state at the lowest level ($d = D$), perturb the mean of its associated Gaussian observation distribution as follows

$$\begin{aligned} \mu_1 &= \mu_0 + u_s \eta \\ \mu_2 &= \mu_0 - u_s \eta \end{aligned}$$

where $\mu_s \sim U[0, 1]$, and η is a simulation parameter that ensures reversibility between split moves and merge moves. When this is a state at $d = 1, \ldots, D - 1$, with more than one children states, split its children into two disjoint sets at random, generate a new sibling state at level d associated with the same parent as the selected state. Update the corresponding multi-level Markov chain parameters accordingly.

Merge two states: Select two sibling states at level d, merge the observation probabilities or the corresponding child-HHMM of these two states, depending on which level they are located in the original HHMM: When $d = D$, merge the Gaussian observation probabilities by making the new mean as the average of the two.

$$\mu_0 = \frac{\mu_1 + \mu_2}{2}, \quad \text{if } |\mu_1 - \mu_2| \leq 2\eta$$

When $d = 1, \ldots, D - 1$, merge the two states by making all the children of these two states the children of the merged state, and modify the multi-level transition probabilities accordingly.

Acceptance Ratio for Different Moves in RJ-MCMC.

The acceptance ratio for Swap simplifies into the posterior ratio because the dimension of the space does not change. Denote Θ as the old model and $\hat{\Theta}$ as the new model:

$$r \triangleq (\text{posterior ratio}) = \frac{P(x|\Theta)}{P(x|\Theta)} = \frac{exp(\widehat{BIC})}{exp(BIC)}.$$

When moves are proposed to a parameter space with different dimension, such as split or merge, we will need two additional terms in evaluating the acceptance ratio: a proposal ratio term to compensate for the probability that the current proposal is actually reached to ensure detailed balance; and a Jacobian term is used to align the two spaces.

Here, $p_{sp}(k)$ and $p_{ms}(k)$ refer to the proposal probabilities, see above, with the extra variable d omitted because split or merge moves do not involve any change across levels.